Journal of Physics: Conference Series

PAPER • OPEN ACCESS

The Geographical Clustering of The Rainfall Stations on Seasonal GSTAR Modeling for Rainfall Forecasting

To cite this article: Alfian Futuhul Hadi *et al* 2018 *J. Phys.: Conf. Ser.* **1028** 012238

IOP Conf. Series: Journal of Physics: Conf. Series **1028** (2018) 012238 doi:10.1088/1742-6596/1028/1/012238

The Geographical Clustering of The Rainfall Stations on Seasonal GSTAR Modeling for Rainfall Forecasting

Alfian Futuhul Hadi[1*], Ira Yudistira[2], Dian Anggraeni[1], and Moh. Hasan[2]

[1]Statistical Laboratory, Department of Mathematics, University of Jember, Jember, 68121, Indonesia
[2]Department of Mathematics, University of Jember, Jember, 68121, Indonesia

*afhadi@unej.ac.id

Abstract. Recent research in time series shows that the data not only have inter-relations with events in the previous time, but also have inter-location linkages. This type of time series data with elements of time and location dependencies are modeled with the Space-Time model. The Space-Time model with heterogeneous research sites is the Generalized Space Time Autoregressive (GSTAR) model. The data that has a seasonal pattern is modeled with seasonal GSTAR by including seasonal elements in the non-seasonal model. In this case, Jember District has 77 rain stations with various regional topography. Based on various characteristics, K-Means cluster analysis is used to obtain optimal rainfall rain stations clusters. This clustering is expected to give better rainfall forecasting result compared with clustering conducted by the Statistics Central of Jember (BPS). The RMSE value can be minimized by including seasonal elements in the model, both in BPS and K-Means clustering. In addition, the K-Means clustering in this study may also reduce the RMSE value of model, both on non-seasonal and seasonal models. The best model for this case is GSTARK-Seasonal (1;1) , ie Seasonal GSTAR model on K-Means clustering.

1. Introduction
Forecasting is a statistical analysis to get an idea of the future development of an event. Forecasting is performed on time series data, which is a sequence of observations taken sequentially based on time with the same interval. Along with its development, research in the field of time series shows that the data encountered not only contain linkages with events in the previous time, but also have inter-location linkages. This conditions are modeled with the Space-Time model. One of the Space-Time models is the STAR [1] model. The STAR model assumes only homogeneous characteristics on its locations. This weakness is then corrected by the Generalized Space Time Autoregressive model (GSTAR) [2].

The GSTAR model is a generalization of the STAR model. Unlike the STAR model, the GSTAR model does not require the same parameter values for all locations, so the GSTAR model is more realistic because in reality more models are found with different parameters for different locations [3] by the form of weighted matrices. The location weights on GSTAR are generally divided into three (uniform location weight, inverse distance and cross-normalization correlation). Similar to the STAR model, the GSTAR model has also been studied in time series modeling, such as rainfall modeling.

In the other side, Jember is one of the districts in Java with a diverse topography of the region. In addition, most of the area is agricultural land and plantation, so that information on rainfall patterns is very useful for farmers to help planning planting patterns. However, research on rainfall forecasting in Jember District is still limited to time series modeling with non-seasonal data. Whereas in reality, rainfall is a periodic occurrence every year, causing a seasonal pattern. In addition, most of the research related to GSTAR model, parameter estimation method used is OLS (Ordinary Least Square). Parameter estimation using OLS on GSTAR model with residual correlation will result in relatively inefficient estimator. This can be corrected with Generalized Least Square (GLS) which is usually used in the Seemingly Unrelated Regression model [4].

Analysis of rainfall forecasting in Jember District in this study using Seasonal GSTAR model with two kinds of data clustering. The first clustering was done based on location of rain station conducted by Central Bureau of Statistics (BPS) Jember, then second clustering was done using K-Means algorithm. The GSTAR model with seasonal elements as well as the K-Means clustering is expected to provide better forecasting results.

2. Experimental Details

2.1. Research Data

The data used in this research is rainfall data from 77 rain stations in Jember District from January 2005 to December 2016. Data is divided into two kinds, namely in-sample data and out-sample data. In-sample data is data used to form forecasting model, that is rainfall data from January 2005 until December 2015. While the out-sample data is data used to check the model's prediction, that is rainfall data from January 2016 until December 2016.

2.2. Data Clustering

77 rain stations in Jember District in this research will be grouped using two kinds of clustering, that is clustering done by Central Bureau of Statistics (BPS) Jember, into four groups, based on the location of the rain station. The second clustering is done by using the Non Hierarchy K-Means clustering algorithm, which based on the coordinates of rain stations, the height of the rainfall station area from the sea level and the intensity of rainfall each month at each rain station. The number of groups in this clustering is four, refers to the clustering conducted by BPS Jember.

The K-Means clustering algorithm is as follows [5]:

a. Determine the number of clusters to be formed.
b. Setting the cluster center Cm, which is the average value of all objects in the mth cluster.
c. Determine the closest distance of each object with each cluster center using the Euclidean distance.

2.3. GSTAR Modeling

The GSTAR modeling is based on the average monthly rainfall data for each region of the two groups. Modeling in each clustering is done twice, Non-Seasonal and Seasonal GSTAR Modeling. So that formed four kinds of GSTAR model.d

2.4. GSTAR Model

If the known series $\{Z(t) : t = 0, \pm 1, \pm 2, \cdots, \pm n\}$ are the multivariate time series of *N* locations, then the Non-Seasonal GSTAR model $(p;\, \lambda_1, \lambda_2, \cdots, \lambda_p)$ in notation matrix as in the following equation [6], so we only added the global model with the seasonal factors.

$$\boldsymbol{Z}(t) = \sum_{k=1}^{p}\left[\emptyset_{k0}\boldsymbol{Z}(t-k) + \sum_{l=1}^{\lambda_k}\emptyset_{kl}\boldsymbol{W}^{(l)}\boldsymbol{Z}(t-1) + \emptyset^{s}{}_{k0}\boldsymbol{Z}(t-s) + \sum_{l=1}^{\lambda_k}\emptyset^{s}{}_{kl}\boldsymbol{W}^{(l)}\boldsymbol{Z}(t-s)\right] + \boldsymbol{e}(t) \quad (1)$$

2.5. Model Order

GSTAR model is a time series forecasting model that depends on the autoregressive order (p) and spatial order (λ_p).. The spatial order of the GSTAR model is restricted to order 1 because higher order models will be difficult to interpret [3]. The autoregressive order is determined by the smallest value of Akaike Information Criterian (AIC) [7]:

$$AIC(i) = \ln(\widehat{\sum_p}) + \frac{2k^2 i}{n} \quad (2)$$

so the order of the model is p such that $AIC(p) = \min_{0 \le i \le p} AIC(i)$.

2.5.1. Inverse Distance Location Weights
Weighted by inverse distance refers to the distance between four locations. The weight of the inverse distance location is in the form of the matrice below [8].

$$W_{ij} = \begin{bmatrix} 0 & \frac{d_{13}+d_{14}}{d_{12}+d_{13}+d_{14}} & \frac{d_{13}+d_{14}}{d_{12}+d_{13}+d_{14}} & \frac{d_{13}+d_{14}}{d_{12}+d_{13}+d_{14}} \\ \frac{d_{23}+d_{24}}{d_{12}+d_{23}+d_{24}} & 0 & \frac{d_{12}+d_{24}}{d_{12}+d_{23}+d_{24}} & \frac{d_{12}+d_{23}}{d_{12}+d_{23}+d_{24}} \\ \frac{d_{23}+d_{34}}{d_{13}+d_{23}+d_{34}} & \frac{d_{13}+d_{24}}{d_{13}+d_{23}+d_{34}} & 0 & \frac{d_{13}+d_{23}}{d_{13}+d_{23}+d_{34}} \\ \frac{d_{24}+d_{34}}{d_{14}+d_{24}+d_{34}} & \frac{d_{14}+d_{34}}{d_{12}+d_{23}+d_{24}} & \frac{d_{14}+d_{24}}{d_{14}+d_{24}+d_{34}} & 0 \end{bmatrix}$$

The above W_{ij} is standardized as Wij^* to obtain $\sum_{i \neq j} Wij^{(l)} = 1$:

Estimation of GSTAR model parameters is done by using GLS (Generalized Least Square) [4] method. Mathematically, the GSTAR model with the order p = 1 and the order $\lambda_p = 1$ is written with GSTAR (11), namely:

$$Z_i(t) = \emptyset_{0i} Z_i(t-1) + \emptyset_{1i} \sum_{j=1}^{N} W_{ij} Z_j(t-1) + e_i(t) \quad (3)$$

With $\emptyset_{ki} = \emptyset_{lk}{}^{(i)}$ for $k = 0, 1$ and $Z_i(t)$ stated observations on $t = 0, 1, \cdots, T$ for location $i = 1, 2, \cdots, N$. Equation (3) above is similar to a linear form $Y_i = X_i \beta_i + \varepsilon_i$, which the residuals on the GSTAR model are correlated between equations or locations. Estimation of parameters by GLS method is done by minimizing the generalized sum of square $\varepsilon' \Omega^{-1} \varepsilon$, which is $\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y$

The best model selection is based on the RMSE (Root Mean Square Error) model. The smaller the RMSE value, the better the model, and vice versa. If M is size of data to be predicted, then the RMSE calculation is formulated with the following equation [9].

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{M} \sum_{t=1}^{M} \left(Z(t) - \hat{Z}(t) \right)^2}$$

3. Results and Discussion

The first clustering is a clustering performed by BPS (Figure 1). JPS BPS grouped 77 rain stations in Jember District into four groups of regions, namely West Jember (⟦Z ⟧_B1) in red, South Jember (⟦Z ⟧_B2) in green, Central Jember (⟦Z ⟧_B3) in blue, and East Jember (⟦Z ⟧_B4) in purple. The clustering is based on the location of the rain station so that the adjacent rain stations are located within a group of territories.

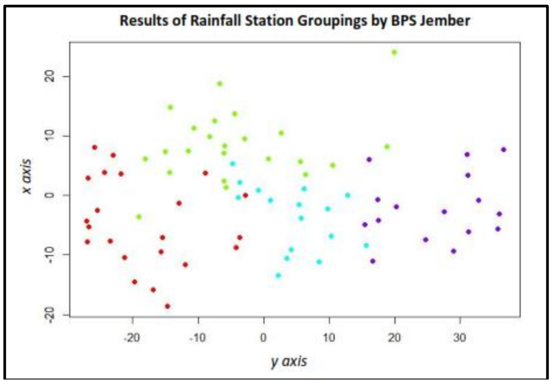

Figure 1. Plot Result of Rain Station Clustering by BPS Jember

The second clustering uses the K-Means algorithm (Figure 2). The number of groups in the K-Means clustering in this study were four, which are region 1 (Z_{K1}) in red, region 2 (Z_{K2}) in green, region 3 (Z_{K3}) in blue, and region 4 (Z_{K4}) in purple. This refers to the clustering conducted by BPS Jember. Based on Figure 2, it appears that several adjacent rain stations are not clustered in one area, but spread in other areas. This is because the clustering is not only based on the coordinates of the rain station but also based on the height of the rainfall station area from the sea level and the intensity of rainfall each month at each rain station.

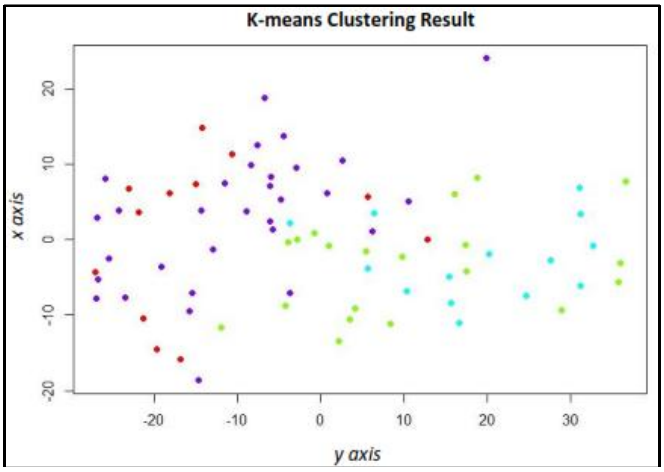

Figure 2. Hadi et al.

3.1. GSTAR Modeling

Modeling in the two clusterings is done twice, Non-Seasonal and Seasonal GSTAR. So that formed four kinds of GSTAR model in this research. AIC model values with the tentative order are presented in Table 1. The smallest AIC values of Non Seasonal GSTAR and Seasonal GSTAR models for BPS clusterings and K-Means clustering are AR 1, which is 32,107; 32,038; 31,095 and 30,508. So it can be concluded that the order p to the four models is 1 and the model formed is $\mathrm{GSTAR_B}(1;1)$ and $\mathrm{Seasonal} - \mathrm{GSTAR_B}(1;1)$ for BPS clusterings, as well as $\mathrm{GSTAR_K}(1;1)$ and $\mathrm{Seasonal} - \mathrm{GSTAR_K}(1;1)$ for K-Means clustering.

Table 1. Summary Model of AIC Value with Tentative Orde

p-th	BPS Clustering	K-Means Clustering

Orde	Non-Seasonal GSTAR	Seasonal GSTAR	Non-Seasonal GSTAR	Seasonal GSTAR
1	32.107	32.038	31.095	30.508
2	32.278	32.250	31.194	30.711
3	32.435	32.453	31.390	30.936
4	32.656	32.693	31.601	31.168
5	32.965	33.033	31.880	31.455
6	33.166	33.310	32.241	31.844
7	33.384	33.539	32.416	32.060

The inverse weight distance is the weight of the location based on the distance between locations. The distances between rainfall station group areas are summarized in Table 2.

Table 2. Distance Among Rainfall Stations Clustering

BPS Clustering				
Region	Z_{B1}	Z_{B2}	Z_{B3}	Z_{B4}
Z_{B1}	0			
Z_{B2}	r_{B1} = 3.31	0		
Z_{B3}	r_{B2} = 1.1	r_{B4} = 2.2	0	
Z_{B4}	r_{B3} = 19	r_{B5} = 3.48	r_{B6} = 2.78	0
K-Means Clustering				
Region	Z_{K1}	Z_{K2}	Z_{K3}	Z_{K4}
Z_{K1}	0			
Z_{K2}	r_{K1} =3.8	0		
Z_{K3}	r_{K2} =2.23	r_{K4} =2.2	0	
Z_{K4}	r_{K3} =1.11	r_{K5} =1.82	r_{K6} =2.2	0

3.2. Parameter Estimation

3.2.1. Parameter Estimation of Non Seasonal GSTAR Model

The results of estimation parameters of Non-Seasonal GSTAR model can be seen in Table 3. There are some non-significant model parameters, it means that these parameters have no effect on the model. Stating that an insignificant parameter can be used to forecast because there is no guarantee that a model formed of a significant parameter will result in a small prediction error value [10].

Based Table 3 and Eq. 3, then the non-seasonal model on BPS clustering is $GSTAR_B(1;1)$

$$\hat{Z}_{B1}(t) = 0.385Z_{B1}(t-1) + 1.373Z_{B2}(t-1) + 1.247Z_{B3}(t-1) + 6.524Z_{B4}(t-1) + e_{B1}(t)$$
$$\hat{Z}_{B2}(t) = -0.171Z_{B1}(t-1) + 0.059Z_{B2}(t-1) - 0.147Z_{B3}(t-1) + 0.182Z_{B4}(t-1) + e_{B2}(t)$$
$$\hat{Z}_{B3}(t) = 0.462Z_{B1}(t-1) + 0.592Z_{B2}(t-1) - 0.758Z_{B3}(t-1) + 0.696Z_{B4}(t-1) + e_{B3}(t)$$
$$\hat{Z}_{B4}(t) = 0.278Z_{B1}(t-1) + 0.080Z_{B2}(t-1) + 0.078Z_{B3}(t-1) + 0.314Z_{B4}(t-1) + e_{B4}(t)$$

and non-seasonal model on K-Means clustering is $GSTAR_K(1;1)$

$$\hat{Z}_{K1}(t) = 0.156Z_{K1}(t-1) + 1.426Z_{K2}(t-1) + 0.970Z_{K3}(t-1) + 0.789Z_{K4}(t-1) + e_{K1}(t)$$
$$\hat{Z}_{K2}(t) = 0.103Z_{K1}(t-1) + 0.105Z_{K2}(t-1) + 0.074Z_{K3}(t-1) + 0.069Z_{K4}(t-1) + e_{K2}(t)$$
$$\hat{Z}_{K3}(t) = -0.764Z_{K1}(t-1) - 0.759Z_{K2}(t-1) + 0.990Z_{K3}(t-1) - 0.759Z_{K4}(t-1) + e_{K3}(t)$$
$$\hat{Z}_{K4}(t) = 0.361Z_{K1}(t-1) + 0.439Z_{K2}(t-1) + 0.496Z_{K3}(t-1) - 0.032Z_{K4}(t-1) + e_{K4}(t)$$

Table 3. Parameter Estimation of Non-Seasonal GSTAR Model

Model	Parameter	Estimation	p-value	Sig
$GSTAR_B(1;1)$	$\emptyset_{B\,10}$	0.385	0.072	.
	$\emptyset_{B\,20}$	0.059	0.030	*
	$\emptyset_{B\,30}$	-0.758	0.003	**
	$\emptyset_{B\,40}$	0.314	0.000	***
	$\emptyset_{B\,11}$	1.188	0.000	***
	$\emptyset_{B\,21}$	-0.110	0.206	
	$\emptyset_{B\,31}$	0.378	0.022	*
	$\emptyset_{B\,41}$	0.069	0.002	**
$GSTAR_K(1;1)$	$\emptyset_{K\,10}$	0.156	0.552	
	$\emptyset_{K\,20}$	0.105	0.009	**
	$\emptyset_{K\,30}$	0.990	0.000	***
	$\emptyset_{K\,40}$	-0.032	0.565	
	$\emptyset_{K\,11}$	0.667	0.001	***
	$\emptyset_{K\,21}$	0.053	0.300	
	$\emptyset_{K\,31}$	-0.507	0.076	.
	$\emptyset_{K\,41}$	0.283	0.000	***

3.2.2. Parameter Estimation of Seasonal GSTAR Model

Table 4 give us the result of parameter estimation of seasonal GSTAR model. Based on Table 4, it can be seen that $Seasonal-GSTAR_B(1;1)$ model, parameter $\emptyset_{B\,20}$, $\emptyset_{B\,40}$, $\emptyset_{B\,31}$, $\emptyset^s_{B\,20}$, $\emptyset^s_{B\,30}$, $\emptyset^s_{B\,40}$, $\emptyset^s_{B\,11}$ and $\emptyset^s_{B\,31}$ are insignificant, while almost all of the parameters on $Seasonal-GSTAR_K(1;1)$ are insignificant, except $\emptyset_{K\,11}$, $\emptyset^s_{K\,10}$ $\emptyset^s_{K\,21}$, dan $\emptyset^s_{K\,41}$.

Table 4. Parameter Estimation of Seasonal GSTAR Model

Model	Parameter	Estimation	p-value	Sig	Model	Parameter	Estimation	p-value	Sig

IOP Conf. Series: Journal of Physics: Conf. Series **1028** (2018) 012238 doi:10.1088/1742-6596/1028/1/012238

Seasonal $-$ GSTAR$_B$ (1;	$\emptyset_{B\,10}$	0.428	0.006	**	Seasonal $-$ GSTAR$_K$(1;	$\emptyset_{K\,10}$	0.234	0.285	
	$\emptyset_{B\,20}$	-0.005	0.412			$\emptyset_{K\,20}$	0.027	0.446	
	$\emptyset_{B\,30}$	0.312	0.045	*		$\emptyset_{K\,30}$	0.129	0.590	
	$\emptyset_{B\,40}$	0.031	0.575			$\emptyset_{K\,40}$	0.060	0.109	
	$\emptyset_{B\,11}$	-0.371	0.078	.		$\emptyset_{K\,11}$	0.579	0.002	**
	$\emptyset_{B\,21}$	0.159	0.001	***		$\emptyset_{K\,21}$	-0.027	0.568	
	$\emptyset_{B\,31}$	-0.179	0.347			$\emptyset_{K\,31}$	0.275	0.180	
	$\emptyset_{B\,41}$	0.131	0.003	**		$\emptyset_{K\,41}$	0.055	0.308	
	$\emptyset^{s}_{B\,10}$	0.758	0.000	***		$\emptyset^{s}_{K\,10}$	0.311	0.062	.
	$\emptyset^{s}_{B\,20}$	-0.091	0.100			$\emptyset^{s}_{K\,20}$	0.047	0.251	
	$\emptyset^{s}_{B\,30}$	0.222	0.230			$\emptyset^{s}_{K\,30}$	0.203	0.213	
	$\emptyset^{s}_{B\,40}$	0.080	0.411			$\emptyset^{s}_{K\,40}$	0.065	0.127	
	$\emptyset^{s}_{B\,11}$	0.020	0.893			$\emptyset^{s}_{K\,11}$	-0.119	0.620	
	$\emptyset^{s}_{B\,21}$	0.064	0.002	**		$\emptyset^{s}_{K\,21}$	0.114	0.036	*
	$\emptyset^{s}_{B\,31}$	0.132	0.372			$\emptyset^{s}_{K\,31}$	-0.057	0.810	
	$\emptyset^{s}_{B\,41}$	0.049	0.014	*		$\emptyset^{s}_{K\,41}$	0.119	0.019	*

Table 4 and Eq. 3 shows that the seasonal model on BPS clustering is $\text{Seasonal} - \text{GSTAR}_B(1;1)$

$$\hat{Z}_{B1}(t) = 0.428Z_{B1}(t-1) - 0.429Z_{B2}(t-1) - 0.390Z_{B3}(t-1) - 2.038Z_{B4}(t-1) + 0.758Z_{B1}(t-12) + 0.023Z_{B2}(t-12) + 0.021Z_{B3}(t-12) + 0.110Z_{B4}(t-12) + e_{B1}(t)$$

$$\hat{Z}_{B2}(t) = 0.247Z_{B1}(t-1) - 0.005Z_{B2}(t-1) + 0.212Z_{B3}(t-1) + 0.263Z_{B4}(t-1) + 0.099Z_{B1}(t-12) - 0.091Z_{B2}(t-12) + 0.085Z_{B3}(t-12) + 0.106Z_{B4}(t-12) + e_{B2}(t)$$

$$\hat{Z}_{B3}(t) = -0.219Z_{B1}(t-1) - 0.028Z_{B2}(t-1) + 0.312Z_{B3}(t-1) - 0.330Z_{B4}(t-1) + 0.161Z_{B1}(t-12) + 0.207Z_{B2}(t-12) + 0.222Z_{B3}(t-12) + 0.243Z_{B4}(t-12) + e_{B3}(t)$$

$$\hat{Z}_{B4}(t) = 0.529Z_{B1}(t-1) + 0152Z_{B2}(t-1) + 0.147Z_{B3}(t-1) + 0.031Z_{B4}(t-1) + 0.198Z_{B1}(t-12) + 0.057Z_{B2}(t-12) + 0.055Z_{B3}(t-12) + 0.080Z_{B4}(t-12) + e_{B4}(t)$$

And the seasonal model on K-means clustering is $\text{Seasonal} - \text{GSTAR}_K(1;1)$

$$\hat{Z}_{K1}(t) = 0.234Z_{K1}(t-1) + 1.238Z_{K2}(t-1) + 0.842Z_{K3}(t-1) + 0.686Z_{K4}(t-1) + 0.311Z_{K1}(t-12) \pm 0.254Z_{K2}(t-12) - 0.173Z_{K3}(t-12) - 0.141Z_{K4}(t-12) + e_{K1}(t)$$

$$\hat{Z}_{K2}(t) = -0.053Z_{K1}(t-1) + 0.027Z_{K2}(t-1) - 0.038Z_{K3}(t-1) - 0.035Z_{K4}(t-1) + 0.222Z_{K1}(t-12) + 0.047Z_{K2}(t-12) + 0.159Z_{K3}(t-12) + 0.149Z_{K4}(t-12) + e_{K2}(t)$$

$$\hat{Z}_{K3}(t) = 0.414Z_{K1}(t-1) + 0.412Z_{K2}(t-1) + 0.129Z_{K3}(t-1) + 0.412Z_{K4}(t-1) - 0.086Z_{K1}(t-12) - 0.085Z_{K2}(t-12) + 0.203Z_{K3}(t-12) - 0.085Z_{K4}(t-12) + e_{K3}(t)$$

$$\hat{Z}_{K4}(t) = 0.070Z_{K1}(t-1) + 0.085Z_{K2}(t-1) + 0.096Z_{K3}(t-1)0.060Z_{K4}(t-1) + 0.152Z_{K1}(t-12) + 0.184Z_{K2}(t-12) + 0.208Z_{K3}(t-12)0.065Z_{K4}(t-12) + e_{K4}(t)$$

3.2.3. *Selection of The Best Model*

The best model is determined by RMSE of each model. The smaller the value of RMSE model, it can be said the better the model. Table 5 shows that the model with the smallest RMSE is $\mathrm{Seasonal} - \mathrm{GSTAR_K}(1;1)$.

Based on the RMSE value in Table 5, it can be seen that incorporating seasonal elements in the model can minimize the RMSE value in both the BPS and the K-Means clustering. In addition, the K-Means clustering in this study may also reduce the RMSE model values, both on non-seasonal and seasonal models. So it can be concluded that the best model is $\mathrm{Seasonal} - \mathrm{GSTAR_K}(1;1)$, which is seasonal-GSTAR on K-Means clustering.

Table 5. RMSE of the Models

Model	RMSE
$\mathrm{GSTAR_B}(1;1)$	163.293
$\mathrm{Seasonal} - \mathrm{GSTAR_B}(1;1)$	155.107
$\mathrm{GSTAR_K}(1;1)$	147.505
$\mathrm{Seasonal} - \mathrm{GSTAR_K}(1;1)$	141.466

3.2.4. *Rainfall Forecasting in the Year of 2017*

The final step in this research is rainfall forecasting in the year of 2017 using the best model that has been obtained, namely $Seasonal - GSTAR_K(1;1)$. Rainfall forecasting in 2017 in four areas of Jember District can be seen in Table 6 and Figure 3. The predicted interval of forecasting results is calculated based on the formula below [11].

$$\mathrm{PI}: P(x_{n+h} \pm z_{\frac{\alpha}{2}}\sigma_n(\mathrm{h})) = 1 - \alpha$$

The black colour graph in Figure 3 shows the average intensity of rainfall from 2005 to 2016, while the red colored graph is the result of rainfall forecasting in 2017. Based on Figure 3, it can be seen that in four areas there is no dry month during 2017. This because the autoregressive order model is 1 so that the data used to predict rainfall of 2017 is rainfall year 2016 $\left(Z(t-1)\right)$, where in 2016 there is no dry month. However, based on the predicted confidence interval in Table 6, it can be seen that there is a possibility of dry months at the locations Z_{K1}, Z_{K3}, and Z_{K4} in July and August.

Table 6. Rainfall Forecasting in The Year of 2017

Location		Z_{K1} (mm^3)		Z_{K2} (mm^3)
Month	Forecast	Confidence Interval	Month	Forecast
January	162	[0 ; 351.55]	284	[0 ; 574.67]
February	271	[162.36 ; 379.64]	455	[362.97 ; 547.03]
March	170	[0 ; 340.71]	310	[95.45 ; 524.55]
April	165	[44.48 ; 285.52]	291	[123.00 ; 459.00]
May	98	[0; 213.72]	175	[16.93 ; 333.07]
June	62	[0 ; 132.36]	148	[36.62 ; 259.38]
July	29	[0 ; 89.55]	81	[0; 182.13]
August	18	[2.72 ; 33.28]	56	[0 ; 117.25]
September	43	[0 ; 132.02]	102	[0 ; 277.41]
October	96	[0 ; 203.17]	160	[0 ; 331.14]
November	157	[19.41 ; 294.59]	329	[142.46 ; 515.54]
December	221	[0 ; 450.38]	344	[47.82 ; 640.18]
Location		Z_{K3} (mm^3)		Z_{K4} (mm^3)
Month	Forecast	Confidence Interval	Month	Forecast
January	254	[0 ; 541.22]	235	[8.08 ; 461.92]
February	353	[232.80 ; 473.20]	320	[223.23 ; 416.77]
March	232	[47.33 ; 416.67]	238	[85.50 ; 390.50]
April	222	[63.49 ; 380.51]	202	[70.30 ; 333.70]
May	127	[0 ; 284.91]	111	[0; 230.64]
June	89	[0 ; 181.24]	84	[9.05 ; 158.95]
July	43	[0; 106.88]	40	[0 ; 103.20]
August	29	[0; 63.46]	24	[5.11 ; 42.89]
September	68	[0 ; 190.99]	60	[0 ; 165.50]
October	129	[0 ; 281.44]	105	[0 ; 223.19]
November	256	[102.09 ; 409.91]	239	[87.78 ; 390.22]
December	292	[30.53 ; 553.47]	257	[1.83 ; 512.17]

The intensity of high rainfall occurred around January-April then continued in November and December. While the lowest rainfall intensity occurred in August. The intensity of rainfall began to decline from April to August and then increased again in September.

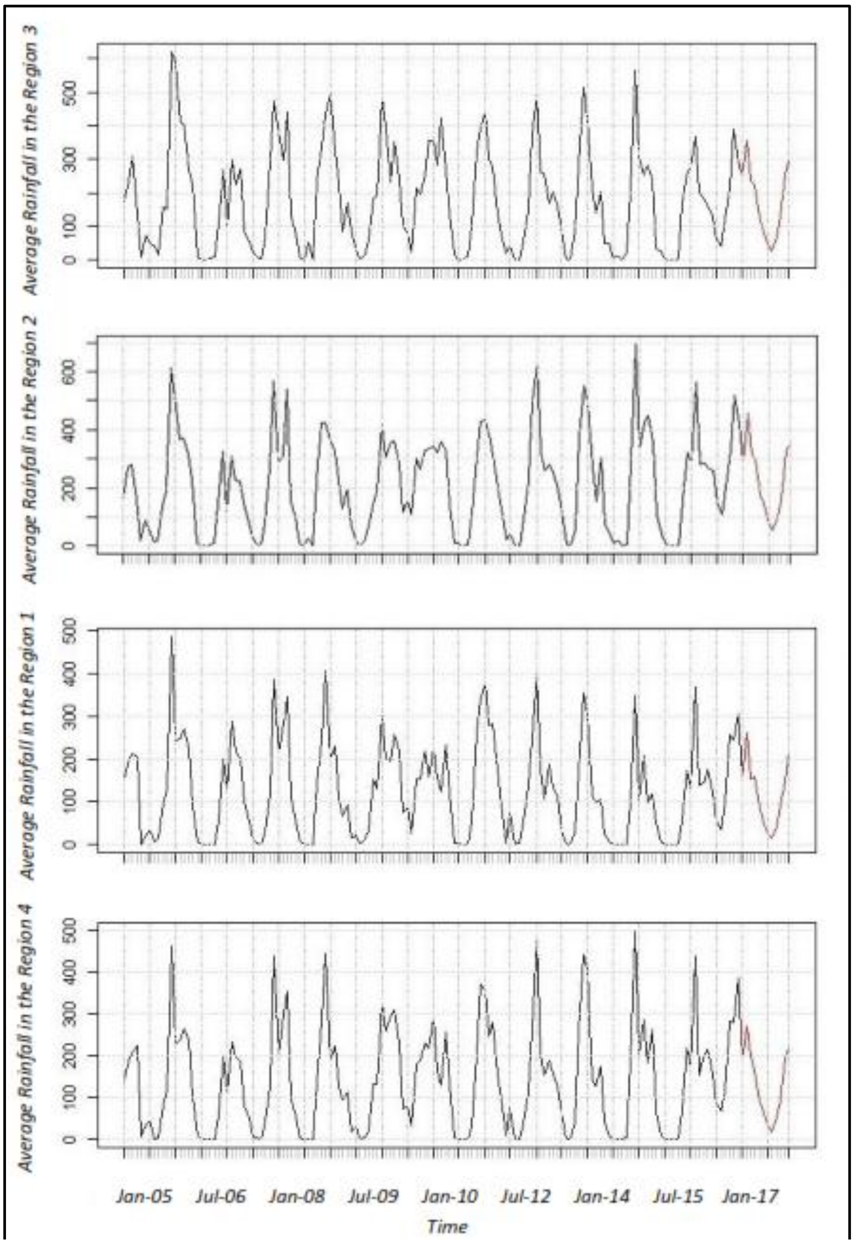

Figure 3. Rainfall in the period of 2005-2016 and Rainfall Forecasting in 2017

4. Conclusion

a. The more optimal clustering of rainfall stations for rainfall forecasting in Jember District is clustering by K-Means Algoritma. This clustering is based on the coordinates of the location of the rain station, the height of the rainfall station area from the sea level, and the intensity of monthly rainfall during 2015.

b. The rainfall forecasting model in Jember District is $\text{Seasonal} - \text{GSTAR}_\text{B}(1;1)$ for rainfall station clustering conducted by BPS and $\text{Seasonal} - \text{GSTAR}_\text{K}(1;1)$. for clustering rain stations using K-Means algorithm. The better model is $\text{Seasonal} - \text{GSTAR}_\text{K}(1;1)$.

c. Based on the model of $\text{Seasonal} - \text{GSTAR}_\text{K}(1;1)$, rainfall forecasting in Jember District inthe year of 2017 in four areas is the highest intensity of rainfall occurred in February, while the lowest rainfall intensity occurred in August. The intensity of rainfall began to decline from March to August and then increased again in September to December (see Table 6 for details).

Acknowledgments
The authors would like to express heartfelt thanks to the Prof. S.-J. Lin and the late Prof. T.-B. Wu from National Tsing-Hua University. C.-S. Peng is grateful to Dr. D.-T. Lee and G.H. Group for their helpful discussions. We also appreciate the Industrial Technology Research Institute of Taiwan (under Contract A351AA5110) and National Science Council of Taiwan (under Contract NSC 101-2120-M-007-003) for providing financial support.

References
[1] Pfeifer P E, and Deutsch S J 1980 A Three Stage Iterative Procedure for Space-Time Modeling. *Technometrics* **22**(1) 397-408
[2] Borovkova S A, Lopuhaa H P, and Ruchjana B N 2002 Generalized S-TAR with Rdanom Weight *Proceeding of the 17th International Workshop on Statistical Modeling* Chania-Greece 139-147
[3] Wutsqa D W, Suhartono, and Sujito B 2010 Generalized Space-Time Autoregressive Modeling. *Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics dan its Applications (ICMSA2010)* Universitas Tunku Abdul Rahman, Kuala Lumpur, Malaysia
[4] Setiawan, Suhartono, and Prastuti M 2016 S GSTAR-SUR for Seasonal Spatio Temporal Data Forecastin, *Malaysian Journal Of Mathematical Sviences* ***10*** 53-65
[5] Agusta Y 2007 K-Means-Penerapan, Permasalahan dan Metode Terkai, Denpasar, Bali: *Jurnal Sistem dan Informatika* **3** 47-60
[6] Borovkova S A, Lopuhaa, and Ruchjana B N 2008 Consistency dan Asymptotic Normality of Least Squares Estimators in Generalized STAR Models, *Journal compilation Statistica Nederldanica*. 487-489
[7] Tsay R S 2005 *Analysis of Financial Time Series* (New Jersey: John Wiley & Sons)
[8] Anggraeni D, Prahutama A, and Andari S 2013 Aplikasi Generalized Space Time Autoregressive (GSTAR) pada Pemodelan Volume Kendaraan Masuk Tol Semarang *Media Statistika* **6**(2) 71-80
[9] Wei W W S 2006 *Time series Analysis : Univariat dan Multivariate Methods* (Canada: Addison Wesley Publishing Company)
[10] Susanti D and Susiswo 2013 *Aplikasi Model GSTAR pada Peramalan Jumlah Kunjungan Wisatawan Empat Lokasi Wisata di Batu* (Malang : Universitas Negeri Malang)
[11] Anderson P L, Meerschaert M M, and Zhang K 2013 Forecasting With Prediction Interval For Periodic ARMA Model *J Time Ser Anal.*, March 1 **34**(2) 187-193